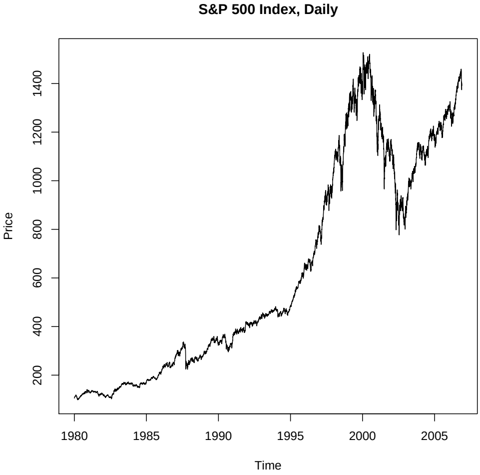
Recent Developments in Measuring Asset Return Volatility and Covariances

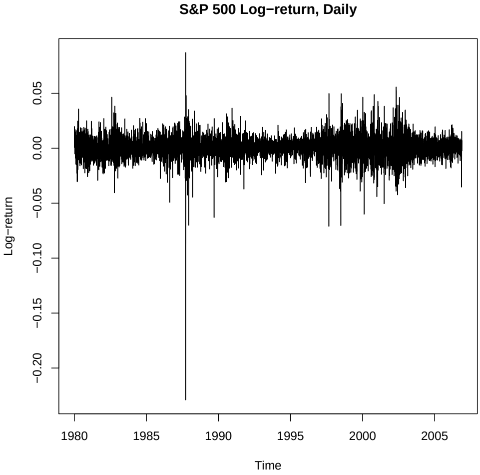
Nikolaus Hautsch
Institute for Statistics and Econometrics
Humboldt-Universität zu Berlin
CASE, CFS, QPL

Valencia, May 5, 2009

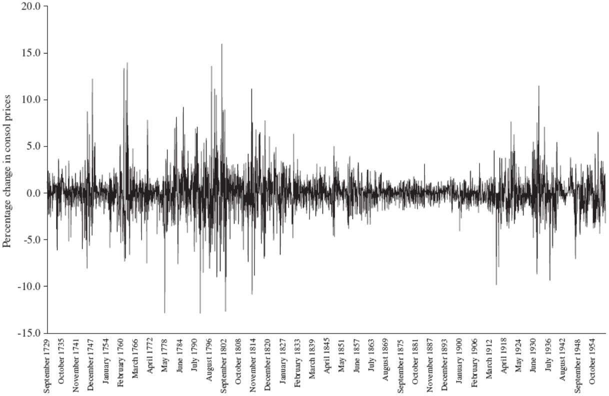
Daily Prices, S&P500, 1980-2007



Daily Returns, S&P500, 1980-2007



U.K. Consol Returns, 1729-1957



Source: Brown, Burdekin and Weidemeier (2006, JFE)

- Volatility changes over time and is clustered in time

Why to Care About Varying Volatility?

Time-varying volatility plays a central role in many areas:

- Sign forecasting and Market Timing
- Default risk
- Risk management
- Asset pricing
- Portfolio allocation
- Hedging
- Option pricing
- Order execution strategies
- ...

Outline

Outline

ARCH

- AutoRegressive Conditional Heteroscedastic (ARCH) model (Engle, 1982, Ecta).
- The ARCH(p) model for log returns r_t is given by

$$
\begin{aligned}
r_t &= \mu_t + \varepsilon_t \\
\varepsilon_t &= z_t \sigma_t, \\
\sigma_t^2 &= \omega + \sum_{j=1}^{p} \alpha_j \varepsilon_{t-j}^2,
\end{aligned}
$$

  where $\mu_t \equiv \mathrm{E}[r_t|\mathcal{F}_{t-1}]$, z_t is an i.i.d. error term with $\mathrm{E}[z_t] = 0$, and $\mathrm{V}[z_t] = 1$.

$\Rightarrow$ Basic Principle: Mean-corrected asset returns $\varepsilon_t = r_t - \mu_t$ are *serially uncorrelated*, but *dependent*.

GARCH

- Problem of the ARCH model: For typical financial time series, often highly parameterized ARCH models are required.
- Generalized ARCH (GARCH), Bollerslev (1986, JoE).
- The GARCH(p,q) model is given by

$$\varepsilon_t = z_t \sigma_t,$$
$$\sigma_t^2 = \omega + \sum_{j=1}^{p} \alpha_j \varepsilon_{t-j}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2,$$

where z_t is an i.i.d. error term with $\mathrm{E}[z_t] = 0$ and $\mathrm{V}[z_t] = 1$.

Properties

- ARMA model for squared (de-meaned) returns:

$$\varepsilon_t^2 = \omega + \sum_{j=1}^{m}(\alpha_j + \beta_j)\varepsilon_{t-j}^2 - \sum_{j=1}^{q}\beta_j \nu_{t-i} + \nu_t$$

with $\nu_t := \varepsilon_t^2 - \sigma_t^2$, $\alpha_j := 0$ for $j > q$, $\beta_j := 0$ for $j > p$ and $m = \max\{p, q\}$.

- Unconditional variance:

$$\mathrm{V}[\varepsilon_t] = \frac{\omega}{1 - \sum_{j=1}^{p}\alpha_j - \sum_{j=1}^{q}\beta_j}.$$

- Kurtosis of ε_t for an GARCH(1,1) process:

$$K_\varepsilon = \frac{3(1 - (\alpha_1 - \beta_1)^2)}{1 - (\alpha_1 + \beta_1)^2 - 2\alpha_1^2} > 3.$$

An (Incomplete) List of GARCH Models ...

- ARCH Engle (1982)
- GARCH Bollerslev (1986)
- IGARCH Bollerslev and Engle (1986)
- Log-GARCH Geweke (1986), Milhøj (1987), Pantula (1986)
- TS-GARCH Taylor (1986), Schwert (1989)
- GARCH-t Bollerslev (1987)
- ARCH-M Engle, Lilien and Robins (1987)
- MGARCH Bollerslev, Engle and Wooldridge (1998)
- CCC GARCH Bollerslev (1990)
- AGARCH Engle (1990)
- CGARCH Engle and Lee (1990)
- EGARCH Nelson (1991)
- SPARCH Engle and Gonzalez-Rivera (1991)
- LARCH Robinson (1991)
- AARCH Bera, Higgins and Lee (1992)
- NGARCH Higgins and Bera (1992)
- QARCH Sentana (1992)
- STARCH Harvey, Ruiz and Sentana (1992)
- TGARCH Zakoian (1994)
- GJR-GARCH Glosten, Jagannathan and Runkle (1993)
- QTARCH Gourieroux and Monfort (1992)

... ctd.

- Weak GARCH Drost and Nijman (1993)
- VGARCH Engle and Lee (1993)
- APARCH Ding, Granger and Engle (1993)
- SWARCH Hamilton and Susmel (1994)
- GQARCH Sentana (1995)
- SGARCH Liu and Brorsen (1995)
- PGARCH Bollerslev and Ghysels (1996)
- HGARCH Hentschel (1995)
- FIGARCH Baillie, Bollerslev and Mikkelsen (1996)
- FIEGARCH Bollerslev and Mikkelsen (1996)
- ATGARCH Crouchy and Rockinger (1997)
- Aug-GARCH Duan (1997)
- STGARCH Gonzalez-Rivera (1998)
- OGARCH Alexander (2001)
- DCC GARCH Engle (2002)
- Flex-GARCH Ledoit, Santa-Clara and Wolf (2003)
- HYGARCH Davidson (2004)
- COGARCH Klüppelberg, Lindner and Maller (2004)

Outline

Stochastic Volatility

- Continuous time random walk for price process $p(t)$:

$$dp(t) = \mu dt + \sigma dW(t),$$

where μ denotes the drift, σ the volatility and $W(t)$ denotes a Brownian motion.

  - Implied by Black-Scholes formula
  - σ constant!

- Time-varying volatility:

$$dp(t) = \mu dt + \sigma(t) dW(t),$$

where $\sigma(t)$ denotes the spot volatility.

- How does $\sigma(t)$ vary over time?

Stochastic Volatility Models

- Cox/Ingersoll/Ross (1985, Etca):

$$\sigma(t) = \eta p(t)^{1/2}$$

- GARCH diffusion (Nelson, 1990, JoE):

$$d\sigma^2(t) = (\alpha - \beta\sigma^2(t))dt + \eta\sigma^2(t)dW(t)$$

- Heston (1993, RFS):

$$d\sigma^2(t) = (\alpha - \beta\sigma^2(t))dt + \eta\sigma(t)dW(t)$$

- Log volatility:

$$d\ln\sigma^2(t) = (\alpha - \beta\ln\sigma^2(t))dt + \eta dW(t)$$

Discrete-Time SV Model

- The SV model by Taylor (1986) is given by

$$r_t = \mu + \sigma_t u_t,$$

$$\ln \sigma_t - \alpha = \phi(\ln \sigma_{t-1} - \alpha) + \eta_t, \quad \eta_t \overset{i.i.d.}{\sim} N(0, \sigma_\eta^2),$$

with $\sigma_\eta^2 = \beta^2(1 - \phi^2)$, $|\phi| < 1$ and u_t and σ_t independent.

- σ_t is log-normally distributed!
- r_t follows a normal-log normal mixture!
- Motivated by mixture-of-distribution hypothesis (Clark, 1973, Ecta).
- Empirically supported?

Estimating Discrete-Time SV Models

- σ_t is a latent process!
- $f(r_t|\mathcal{F}_{t-1})$ is not available in closed form!
- Different ways to estimate the model:
  - GMM, Melino and Turnbull (1990, JoE),
  - QML estimation, Harvey, Ruiz & Shephard (1994, RES)
  - Eff. method of moments, Gallant, Hsie & Tauchen (1997, JoE)
  - Simulated maximum likelihood, Danielson (1994, JoE),
  - Eff. import. sampling, Liesenfeld & Richard (2003, JEmpF),
  - Markov chain Monte Carlo (MCMC), Kim, Shephard & Chib (1998, RES), Hautsch & Ou (2008, Appl.Quant.Fin., Springer)

Outline

Integrated Variation

- Assume the diffusion

$$dp(t) = \mu(t)dt + \sigma(t)dW(t).$$

- Goal: Estimate the variance over the (normalized) interval $[0, 1]$ (representing e.g. a trading day).
- Then, the variance of $p(1) - p(0)$ given the volatility sample path $\{\sigma(\tau), 0 \leq \tau \leq 1\}$ is computed as

$$IV \equiv \int_0^1 \sigma^2(\tau)d\tau$$

- *IV* is called *integrated variation* or *integrated volatility*.

Realized Volatility

- How to measure the integrated variation

$$IV = \int_0^1 \sigma^2(\tau) d\tau \quad ?$$

- Intuitive: Sum of m squared returns of length $\Delta = m^{-1}$

$$RV^m \equiv \sum_{j=1}^{m} (p(j\Delta) - p((j-1)\Delta))^2 \equiv \sum_{j=1}^{m} r_{j\Delta,m}^2.$$

- RV is called *realized variance* or *realized volatility*.

Theory of Quadratic Variation

- Barndorff-Nielsen and Shephard (2002, JRRS B):

$$\sqrt{m}(RV^m - IV)|IQ \xrightarrow{d} N(0, 2\Delta IQ)$$

- The *integrated quarticity IQ* can be consistently estimated using the *realized quarticity*:

$$RQ^m \equiv \frac{1}{3\Delta}\sum_{j=1}^{m} r_{j\Delta,m}^4 \quad \xrightarrow{p} \quad IQ \equiv \int_0^1 \sigma^4(\tau)d\tau$$

- Then:

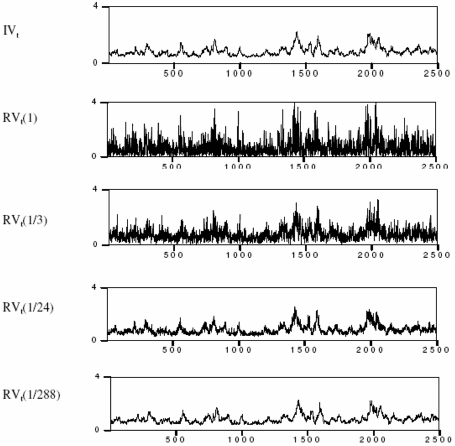
$$\sqrt{m}\frac{RV^m - IV}{\sqrt{2RQ^m}} \overset{a}{\sim} N(0,1).$$

Implications

- Asymptotic variance declines with increasing $m = \Delta^{-1}$!
- 'In-fill' asymptotics: Sampling on highest possible frequencies crucial!

- Measuring the realized volatility over non-trivial intervals avoids double asymptotics required for estimating $\sigma(t)$.
- Completely model-free measure!

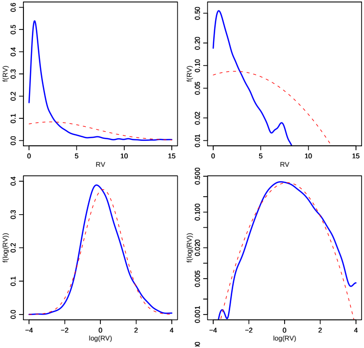
Different Realized Volatility Estimators



Source: Tim Bollerslev, CASE-QPL Lecture 2009, Berlin

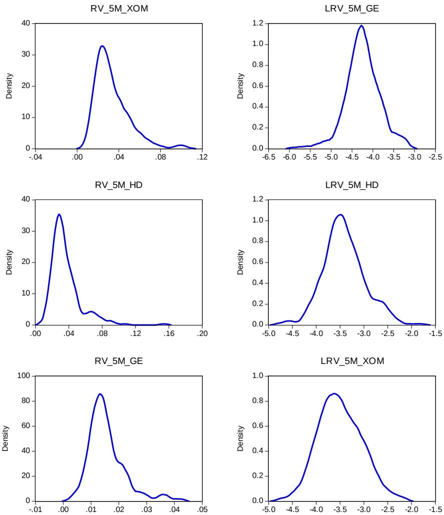
Empirical Properties of the RV Estimator

- The unconditional distribution of realized volatility is approximately log-normal

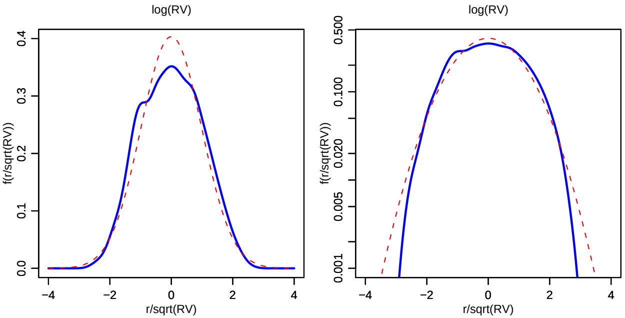


Source: Härdle, Hautsch & Pigorsch (2008, Appl.Quant.Fin., Springer)

RVs for XOM, HD and GE, NYSE, 2006

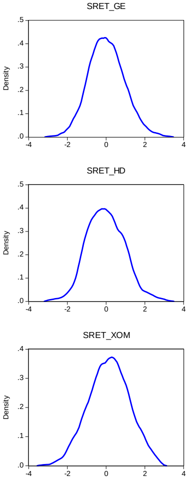


The Distribution of $r/RV^{1/2}$

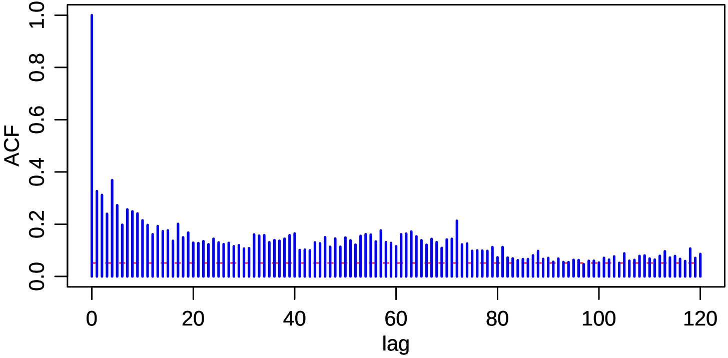


Source: Härdle, Hautsch & Pigorsch (2008, Appl.Quant.Fin., Springer)

$r/RV^{1/2}$ for XOM, HD and GE, NYSE, 2006

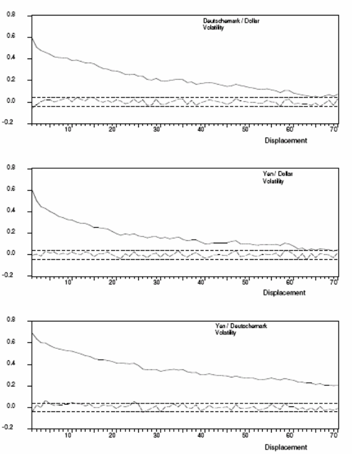


Empirical ACF of RV, IBM, 2001-2006



Source: Härdle, Hautsch & Pigorsch (2008, Appl.Quant.Fin., Springer)

Empirical ACFs of Realized Volatility



Source: Andersen et al (2001, JASA)

Realized Volatility Reveals Long Range Dependence

- If $RV_t \sim I(d)$, for $t = 1, 2, \ldots$, denoting days, then

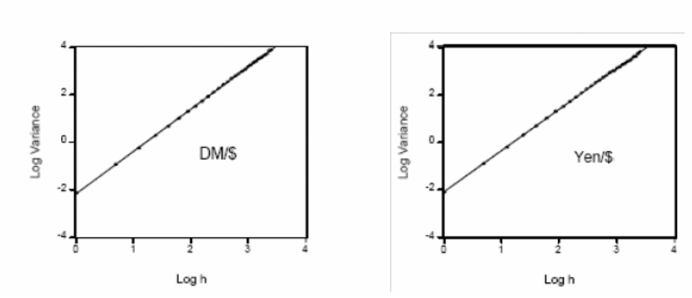
$$\mathrm{V}\left[\frac{1}{h}\sum_{j=1}^{h} RV_{t+j}\right] \approx ch^{-\alpha}$$

with $\alpha = 2H - 2 = 1 - 2d$.

- Consequently:

$$\mathrm{V}\left[\sum_{j=1}^{h} RV_{t+j}\right] \approx ch^{2d+1}$$

Realized Volatility Reveals Long Range Dependence



Source: Andersen et al (2001, JASA)

Implications

- The distribution of realized volatility is approximately log-normal.
- Realized volatility is fractionally integrated.
- RV-standardized returns are approximately normal.

- Distribution of returns is approximately lognormal-normal as advocated in Taylor's (1986) SV model!

Modelling Implications

- Dynamical properties suggest using long memory models.
- Fractionally integrated AR model (Andersen et al, 2003, Ecta):

$$\Phi(L)(1-L)^d(\ln RV_t - \mu) = \varepsilon_t,$$

where $\Phi(L) = 1 - \sum_{j=1}^{p} \alpha_j L^j$.

Outline

Market Microstructure Frictions

- Problem in practice: Market microstructure frictions!
- In reality, we can only observe

$$p(t) = p^*(t) + u(t),$$

  where
  - $p(t)$: observed (log) price,
  - $p^*(t)$: "efficient" (fundamental) latent (log) price,
  - $u(t) \sim WN$ captures so-called market microstructure "noise" (bid-ask spread, price discreteness ...).
- $p^*(t)$ is assumed to follow the diffusion

$$dp^*(t) = \mu(t)dt + \sigma(t)dW(t).$$

- Under the presence of market microstructure noise:

$$r_{j\Delta,m} \equiv p(j\Delta) - p((j-1)\Delta) \equiv r^*_{j\Delta,m} + \varepsilon_{j\Delta,m}$$

with

$$\begin{aligned} r^*_{j\Delta,m} &\equiv p^*(j\Delta) - p^*((j-1)\Delta) \\ \varepsilon_{j\Delta,m} &\equiv u(j\Delta) - u((j-1)\Delta). \end{aligned}$$

- Then, $\lim_{\Delta \to 0} \mathrm{E}[(r^*_{j\Delta,m})^2] = 0$ while

$$\lim_{\Delta \to 0} \mathrm{E}[\varepsilon^2_{j\Delta,m}] > 0!$$

- Noise term dominates for $\Delta \to 0$!

Dynamic Implications of Noise

- If $u(t)$ is i.i.d.,
  - $\varepsilon_{j\Delta,m}$ follows an MA(1) process.
  - Observed returns $r_{j\Delta,m}$ are first-order (negatively) autocorrelated.
- If $u(t)$ is autocorrelated, observed returns $r_{j\Delta,m}$ are high-order autocorrelated.
- Suggests HAC-type estimators!

What Can We Do?

- Just ignoring noise?
- Sparsely Sampling at Some Lower Frequency?
- Sparsely Sampling at Some Optimal Frequency?
- Bias Corrections?
- Modelling the Noise Parametrically?
- Subsampling and Averaging?
- Pre-Averaging?
- Kernel estimators?

RV Estimation Under the Presence of Noise

- Hansen & Lunde (2006, JBES), Bandi & Russell (2006, JBES)
- Assume, $u(t)$, is a zero mean stationary process with $\pi(s) \equiv \mathrm{E}[u(t)u(t+s)]$.
- Then, it can be shown that

$$\mathrm{E}[RV^m] = 2\rho_m + 2m[\pi(0) - \pi(\Delta)],$$

where $\rho_m \equiv \mathrm{E}[\sum_{i=1}^m r^*_{i\Delta,m}\varepsilon_{i\Delta,m}]$.

- Asymptotic bias:

$$\lim_{m\to\infty} \mathrm{E}[RV^m - IV] = 2\rho - 2\pi'(0),$$

with $\rho \equiv \lim_{m\to\infty} \mathrm{E}[\sum_{i=1}^m r^*_{i\Delta,m}\varepsilon_{i\Delta,m}]$.

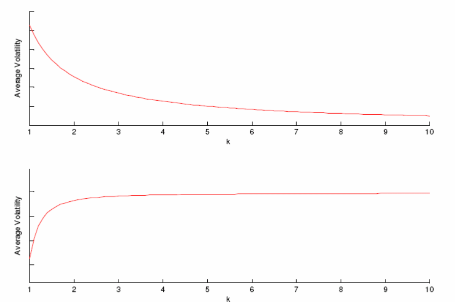
- For independent noise, we have $\pi'(0) = -\infty$ and thus:

$$\mathrm{E}[RV^m] = IV + 2m\omega^2.$$

- For large m, the RV diverges to infinity linearly in m!
- $RV(2m)^{-1}$ consistently estimates the noise variance ω^2!
- Market microstructure noise totally swamps the variance of the price signal!

Sparsely Sampling?

- *Volatility signature plots*: Plot the sample mean of RV^m, $t = 1, 2, \dots, T$ against Δ



Representative volatility signature for liquid and non-liquid assets, k: sampling frequency in minutes.

Source: Andersen, Bollerslev, Diebold and Labys (1999, Risk)

Bias Correction: A Simple HAC Estimator

- Zhou(1996, JBES):

$$RV_Z^m \equiv \sum_{i=1}^{m} r_{i\Delta,m}^2 + 2\sum_{i=1}^{m} r_{i\Delta,m} r_{(i-1)\Delta,m}$$

  - Unbiased!
  - Inconsistent: Asymptotic variance increasing in m!
  - In the absence of noise, $\mathrm{V}[RV_Z] > 3\mathrm{V}[RV]$ (approx.)!
  - Optimal sampling frequencies based on MSE minimization.

Parametric Modelling of Noise

- Ait-Sahalia, Mykland and Zhang (2005, RFS)
- Assume the following (intraday) discrete (log) price process

$$p_{j\Delta} = p^*_{j\Delta} + u_{j\Delta}, \quad u_{j\Delta} \overset{i.i.d.}{\sim} (0, \omega^2),$$

  where

$$r^*_{j\Delta,m} := p^*_{j\Delta} - p^*_{(j-1)\Delta} \overset{i.i.d.}{\sim} (0, \sigma^2\Delta)$$

  with $r^*_{j\Delta,m}$ independent of $u_{j\Delta}$.

- Then, $r_{j\Delta,m} = r^*_{j\Delta,m} + u_{j\Delta} - u_{(j-1)\Delta}$ can be re-parameterized as a MA(1) process,

$$r_{j\Delta,m} := \mu_{j\Delta,m} + \eta\mu_{(j-1)\Delta,m},$$

  where $\mu_{j\Delta,m} \sim (0, \gamma^2)$.

- The parameters η and γ^2 can be identified by

$$\gamma^2(1+\eta^2) = \mathrm{V}[r_{j\Delta,m}] = \sigma^2\Delta + 2\omega^2$$
$$\gamma^2\eta = \mathrm{Cov}[r_{j\Delta,m}, r_{(j-1)\Delta,m}] = -\omega^2$$

- Consequently, the daily variance σ^2 as well as the microstructure noise can be estimated by

$$\widehat{\sigma}^2 = \widehat{IV} = \Delta^{-1}\widehat{\gamma}^2(1-\widehat{\eta})^2,$$
$$\widehat{\omega}^2 = -\widehat{\gamma}^2\widehat{\eta}$$

where $\widehat{\gamma}^2$ and $\widehat{\eta}$ are ML estimates based on a MA(1) process using high-frequency returns.

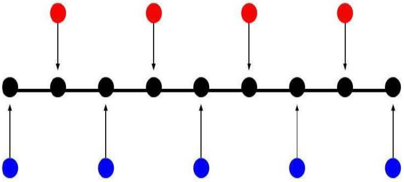
Optimal Sampling Frequencies

- Hansen & Lunde (2004, JBES), Bandi & Russell (2006, JFEC)
- Define $\lambda \equiv \omega^2/IV$ and let $t_{0,m}, \dots, t_{m,m}$ be such that $\mathrm{V}[r^*_{i\Delta,m}] = IV/m$ (business time sampling).
- Then, the optimal sampling frequencies for RV^m and RV^m_Z are given by

$$m_0^* \approx (2\lambda)^{-2/3}$$
$$m_1^* \approx \sqrt{3}(2\lambda)^{-1}$$

Sub-Sampling

- Sparse (even if optimal) sampling discards a lot of data.
- Sub-Sampling: average of the estimates calculated over different sample sets
- Introduced by Zhang, Mykland and Ait-Sahalia (2005, JASA)
- Idea: Divide the time domain grid into K subgrids of size $\overline{m}$.



- *Subsampling* is the average of the estimates calculated over the different subgrids.

$$RV^{(avg)} = \frac{1}{K} \sum_{k=1}^{K} RV^{k,\overline{m}}$$

- Bias correction: $RV_{ZMA} = RV^{(avg)} - RV^{all}$, where '*all*' is associated with sampling over all observations.
- Extension: Multi-scale estimator, Zhang (2006, Bnlli)

Realized Kernels

- If the returns $r_{j\Delta,m}$ are autocorrelated, then

$$\sum_{i=1}^{m} r_{i\Delta,m}^2 \xrightarrow{p} \sum_{j=-m}^{m} \gamma_j,$$

where

$$\gamma_j \equiv \mathrm{E}[r_{i\Delta,m} r_{(i-j)\Delta,m}].$$

- Kernel-based estimators:

$$K(r) = \gamma_0 + \sum_{j=1}^{H} 2k\left(\frac{j-1}{H}\right)\gamma_j.$$

Realized Kernel Estimator

- Modified Tukey-Hannig kernel (Barndorff-Nielsen et al (2008, Ecta):

$$k_{TH}(x) = \left\{1 - \cos \pi (1-x)^2\right\}/2,$$

  with bandwidth $H^* = c^* \sqrt{m}$ and c^* chosen optimally in dependence of ω, IV and IQ.
- Optimal rate of convergence $m^{1/4}$.
- Requires pre-estimates of ω, IV and IQ, e.g. based on Ait-Sahalia, Mykland and Zhang (2005, RFS) ML estimator.

Other Estimators

- Pre-averaging estimator, Jacod, Li, Mykland, Podolskij and Vetter (2007, WP), Hautsch and Podolskij (2009, WP)
- Business-time sampling, Oomen (2006, JBES)
- Alternation estimator, Large (2005, WP)
- ...

Accounting for jumps in the price process:

- Realized bipower variation estimator, Barndorff-Nielsen and Shephard (2004, JFEC)
- Range-based estimators, Christensen and Podolskij (2007, JoE), Martens and van Dijk (2007, JoE)
- ...

Outline

Quadratic Covariation

- p-dimensional log price process (asynchronously) observed over $[0, 1]$:

$$X(t) = (X^{(1)}(t), X^{(2)}(t), \dots, X^{(p)}(t))'$$

- Observation times for the $i-th$ asset: $t_1^{(i)}, t_2^{(i)}, \dots$
- Efficient price process $Y(t)$ follows Brownian semimartingale

$$Y(t) = \int_0^t a(u)du + \int_0^t \sigma(u)dW(u),$$

where a is a predictable locally bounded drift process, σ is a càdlàg volatility matrix process and W is a vector of independent Brownian motions.

- Quadratic covariation of Y:

$$[Y] = \int_0^1 \Sigma(u)du, \qquad \text{where} \qquad \Sigma = \sigma\sigma'$$

and

$$[Y] = \operatorname*{plim}_{m\to\infty} \sum_{j=1}^{m} \{Y(t_j) - Y(t_{j-1})\}\{Y(t_j) - Y(t_{j-1})\}'.$$

- Market microstructure effects:

$$U_j^{(i)} = X(t_j^{(i)}) - Y(t_j^{(i)}), \qquad j = 0, 1, \ldots, m^{(i)}.$$

- Noise process $U_j^{(i)}$ is covariance stationary with
  (i) $\mathrm{E}[U_j^{(i)}] = 0$, and
  (ii) $\sum_h |h\Omega_h| < \infty$, where $\Omega_h = \mathrm{Cov}[U_j, U_{j-h}]$.

Realized Covariance

- Realized *covariance*:

$$RCov^m \equiv \sum_{j=1}^{m} r_{j\Delta,m} r'_{j\Delta,m}$$

where $r_{j\Delta,m} \equiv X(j\Delta) - X((j-1)\Delta)$, $j = 1, \ldots, m = \Delta^{-1}$.

- If $X = Y$ (i.e. U=0),

$$RCov^m \xrightarrow{p} \int_0^1 \Sigma(s) ds$$

for $m \to \infty$.

- Under the absence of noise, $RCov^m$ is consistent and asymptotically normal (Barndorff-Nielsen and Shephard, 2004, Ecta).

Realized Correlations

- The realized correlation between asset i and j is given by

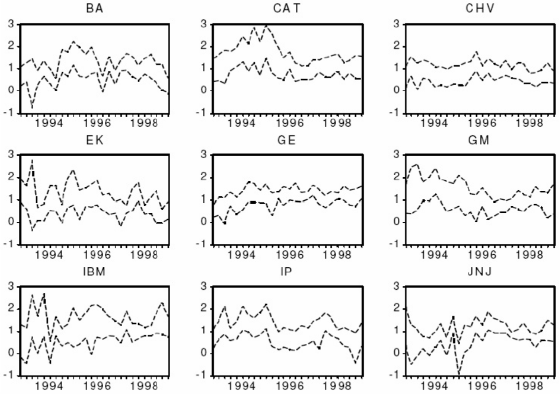
$$RCorr_{ij}^{m} = \frac{RCov_{ij}^{m}}{\left\{RCov_{ij}^{m}\right\}^{1/2}\left\{RCov_{ij}^{m}\right\}^{1/2}}$$

- Realized Betas:

$$R\beta_i = \frac{\left\{RCov_{ip}^{m}\right\}}{\left\{RCov_{pp}^{m}\right\}},$$

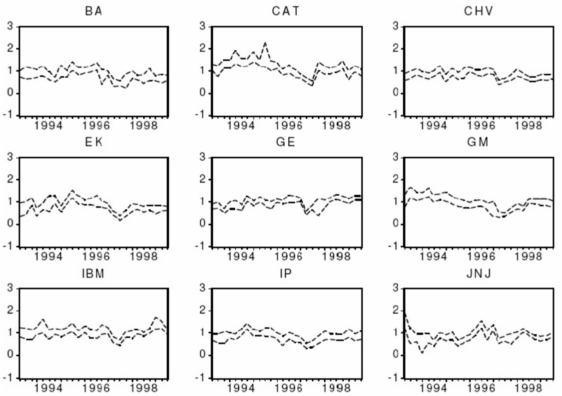
where $\left\{RCov_{ip}^{m}\right\}$ denote the realized covariance between asset i and the market portfolio p.

95% Confidence Intervals for Quarterly Realized Betas Based on Daily Returns for Dow Jones Stocks, 1993-1999



Source: Barndorff-Nielsen and Shephard (2004, Ecta)

95% Confidence Intervals for Quarterly Realized Betas Based on 15min Returns for Dow Jones Stocks, 1993-1999



Source: Barndorff-Nielsen and Shepard (2004, Ecta)

Challenges in Covariation Estimation

- Positive definiteness (invertibility?)
- Well-conditioned (inversions numerically stable?)
- Efficiency (not throwing away too much data due to sparse sampling)
- Market microstructure effects
- Asynchronicity of observations in time, Epps (1979, JASA)

Hayashi-Yoshida Estimator

- Hayashi and Yoshida (2005, Bnlli): Handling the asynchronicity
- Denote $\Pi^A = \{t_i\}_{i=0,1,2,\dots,M_A}$ and $\Pi^B = \{t_j\}_{j=0,1,2,\dots,M_B}$ to be the sets of observation times for two processes A and B.
- Pairwise estimator based on the sum of all overlapping returns:

$$HY = \sum_{i=1}^{M_A}\sum_{j=1}^{M_B} r_A(I^i) r_B(J^j) 1_{\{I^i \bigcap J^j \neq \emptyset\}},$$

where $I^i \equiv (t_{i-1}, t_i)$ and $J^j \equiv (t_{j-1}, t_j)$.

Properties of the Hayashi-Yoshida Estimator

- Positive definite.
- Only applicable to bivariate processes. Pairwise estimation problematic in high dimensions!
- Unbiased and consistent under the absence of noise.
- Bias-correction in the presence of noise: Griffin & Oomen (2006, WP), Voev & Lunde (2007, JFEC)

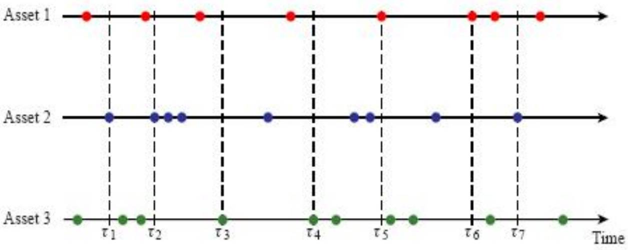
Multivariate Realized Kernel Estimator

- Barndorff-Nielsen, Hansen, Lunde and Shephard (2008, WP)

Modelling strategy:

- Synchronizing the data: Refresh time sampling
- Accounting for noise: realized (multivariate) kernel
- Positive semi-definiteness induced by choice of kernel
- Optimal bandwidth selection based on signal-to-noise ratio

Refresh Time Sampling



- Sample whenever all asset prices have been updated ('refreshed').
- Induces transaction time synchronization.
- Results into n non-overlapping intervals.

Multivariate Realized (Parzen) Kernel

Synchronized returns at time j: $r_j = [r_{1,j}, r_{2,j}, ..., r_{p,j}]$ for p assets.

$$K(X) = \sum_{h=-n+1}^{n-1} k\left(\frac{h}{H+1}\right)\Gamma_h,$$

where

$$\Gamma_h = \begin{cases} \sum_{j=h+1}^{n} r_j x_{j-h}^{'} & \text{for} \quad h \geq 0 \\ \sum_{j=-h+1}^{n} r_{j+h} x_j^{'} & \text{for} \quad h < 0 \end{cases}$$

and

$$k(x) = \begin{cases} 1 - 6x^2 + 6x^3 & 0 \leq x \leq 1/2 \\ 2(1-x)^3 & 1/2 \leq x \leq 1 \\ 0 & x > 1 \end{cases}$$

Outline

Blocking Multivariate Realized Kernels

- Hautsch, Kyj and Oomen (2009, WP)

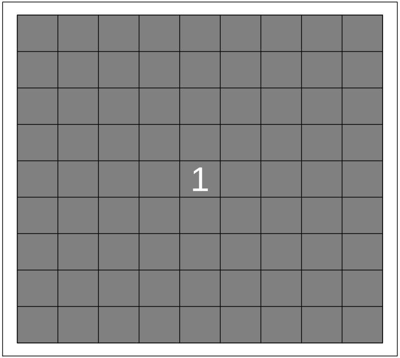
Motivation: Refresh Time Sampling makes

- inefficient use of the data (dramatic if p is high!)
- covariance estimator dependent on 'slowest' assets

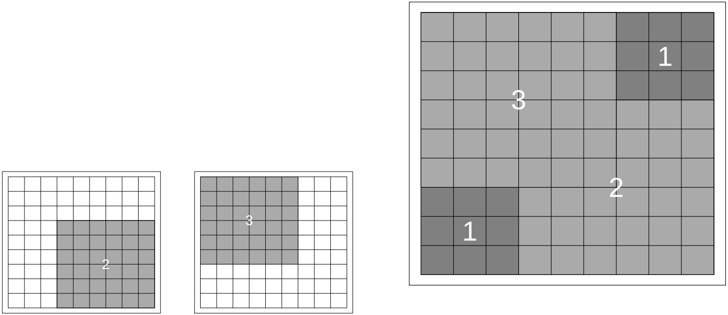
Idea:

- Blocking: limit data reduction due to Refresh Time Sampling
  - Group similar assets, in terms of trading frequency, into blocks
  - Identifying groups based on mixture models
- Regularization: obtaining invertible and numerically well-conditioned estimator
  - Random Matrix Theory

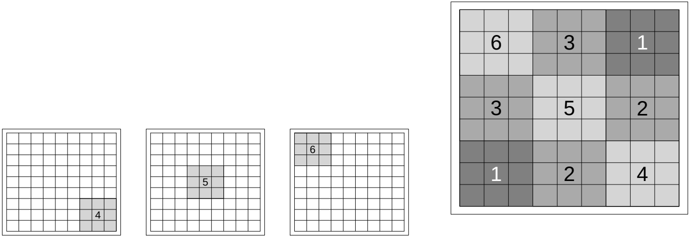
Blocking: Step 1



Blocking: Step 2

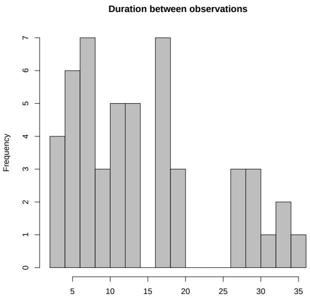


BLocking: Step 3



Block Size Determination

- Group according to liquidity characteristics (e.g. # of trades)
- Isolate similar assets into homogeneous groups (non-synchroneity)
- Grouping: motivated by estimator



Method:

- Finite Mixture Models: Fraley and Raftery (2002)

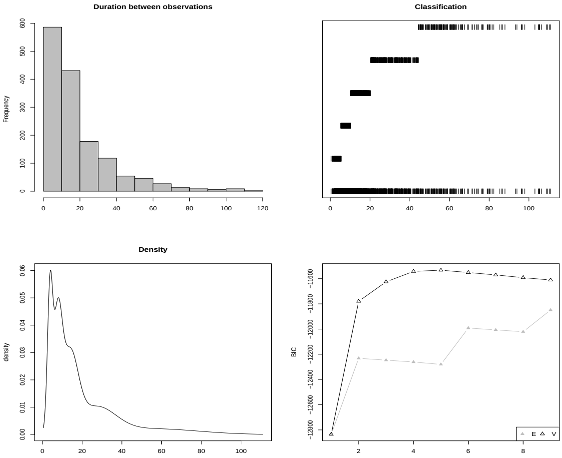
Finite Mixture Models

1. Mixture models

$$L_{MIX}(\theta_1, ..., \theta_G; \tau_1, ...\tau_G|y) = \prod_{i=1}^{n} \sum_{k=1}^{G} \tau_k f_k(y_i|\theta_k),$$

2. EM algorithm for MLE for mixture models
   - ▷ unobserved portion of data is group assignment
3. Model selection:
   - ▷ BIC determines the number of groups

Finite Mixture Model Example



Regularization

Implications of Blocking:

(+) less data reduction due to refresh time sampling

(-) lose positive semi-definiteness of original realized kernel

Perspective:

- Applications call for:
  1. **positive definite** covariance matrices
  2. **well-conditioned** covariance matrices
- Regularize via Random Matrix Theory
  - ▷ directly address negative/vanishing eigenvalue problem

Eigenvalue Projection

- Let C being the estimated correlation matrix with spectral decomposition $C = QLQ^{'}$, where $L = diag(l_i)$ and Q is orthogonal.
- Then, project C on the positive semi-definite cone by setting all negative eigenvalues equal to zero:

$$C_{+} := Qdiag(max(l_i, 0))Q^{'}.$$

- Guarantees positive-definitess but not well-conditioning.

Eigenvalue Cleaning

- Clean for noisy eigenvalues inducing ill-conditioning.
- Identify noisy eigenvalues by comparing the correlation matrix with the identity matrix (inducing independence).
- If $p \to \infty$, $m \to \infty$ and $Q = m/p \geq 1$, the Marchenko Pastur pdf of the eigenvalues $\rho_C(\lambda)$ is given as

$$\rho_C(\lambda) = \frac{Q}{2\pi\sigma^2} \frac{\sqrt{(\lambda_{max} - \lambda)(\lambda - \lambda_{min})}}{\lambda},$$

where

$$\sigma^2 \lambda_{min}^{max} = (1 + 1/Q \pm 2\sqrt{1/Q}),$$

with $\lambda \in [\lambda_{min}, \lambda_{max}]$, and σ^2 being equal to the variance elements of C.

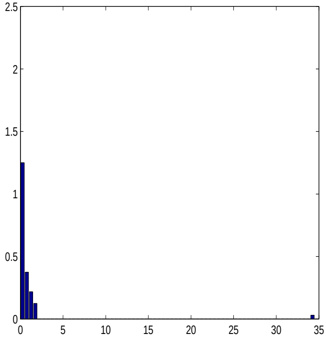
- Remove eigenvalues with $l_i > \lambda_{max}$ (associated with strong common components, "signals").
- (Re-)Compute the remaining contribution of the total variance: $\sigma^2 = 1 - l_1/p$.
  - "Tightening" of the Marchenko Pastor pdf
  - Allows for smaller signals to be better identified
- Recompute λ_{max} and repeat the steps until maximal fit is achieved.

- Then, the regularized correlation matrix is obtained by

$$\widetilde{C} = diag(\widehat{C}^{-1/2})\widehat{C}diag(\widehat{C}^{-1/2}),$$
$$\widehat{C} = Q\widehat{L}Q^{'},$$
$$\widehat{L} = diag(\widehat{l_i}),$$
$$\widehat{l_i} = \begin{cases} l_i & \text{if} \quad l_i > \lambda_{max} \\ \frac{trace(C_+) - \sum_{(l_i > \lambda_{max})} l_i}{p - (\text{No. of } l_i > \lambda_{max})} & \text{otherwise} \end{cases}$$

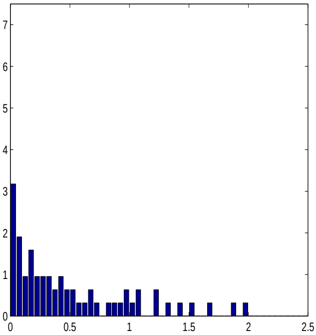
- Smallest eigenvalues are inflated, signals remain unchanged!

Eigenvalue Cleaning I



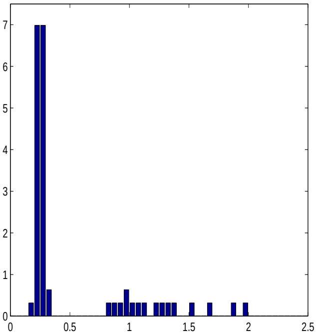
- Estimates eigenvalues for matrix of dim=64
- Clearly see 'market' factor seperate from the rest > 30
- Want to extract some more signals from the 'bulk' and regularize the noise in such a way that it is 'harmless'

Eigenvalue Cleaning II



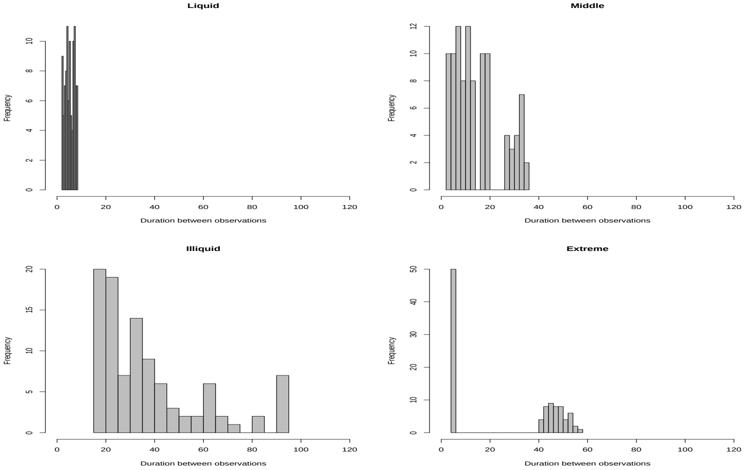
- We zoom in on the bulk
- RMT tests against the null hypothesis that the matrix is uncorrelated, independent assets
- Eigenvalues greater than λ_{max} are regarded as rejecting the null, and the rest are assumed to be noise
- What should we do with 'noise'?

Eigenvalue Cleaning III



- Eliminate negative eigenvalues by projecting onto positive semi-definite cone
- Vanishing eigenvalues result in ill-conditioning (numerical problems)
- Condition by replacing the distribution of 'noise' with the mean 'noise'
- Small eigenvalues are inflated

Observation Frequencies for SP1500



Simulation: Design based on SP1500

Market Microstructure Effects: $X_j = Y_j + U_j, \quad U_j \sim N(0, \omega^2)$

	Noise Ratio $\gamma = \frac{\omega^2}{IV/m}$				
	Q5	Q25	Q50	Q75	Q95
Bottom 600	0.22	0.27	0.34	0.41	0.63
Middle 400	0.23	0.31	0.38	0.46	0.76
Top 500	0.20	0.29	0.36	0.46	0.94

Consider: $\gamma \in (0.25, 0.375, 0.5, 1)$

Simulation: Estimators

1. RK: Realized Kernel estimator.
2. BLOCK: Blocked estimator.
   - Five blocks of equal size
3. RMTBLOCK: BLOCK regularized via Random Matrix Theory
   - Five blocks of equal size

Evaluation Criteria

Let $\widehat{\Sigma}_t$ be our estimate of Σ_t and $\widehat{\lambda}$ be the estimates of the eigenvalues λ.

(i) Check positive semi-definiteness.

$$PSD = \begin{cases} 1 & \text{if} \quad \widehat{\lambda}_{min} > 0 \\ 0 & \text{otherwise} \end{cases}$$

(ii) Evaluate distance from 'true' covariance: Scaled Frobenius norm

$$\|\widehat{\Sigma}_t - \Sigma_t\|_{F_p} = (trace(AA^T)/p)^{1/2},$$

where $A = \widehat{\Sigma}_t - \Sigma_t$.

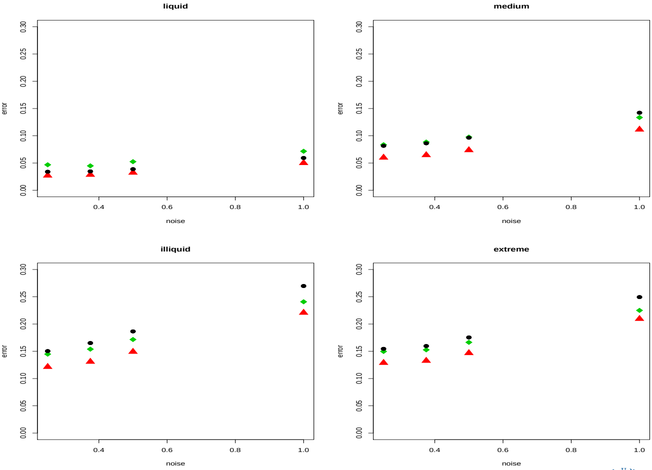
Positive Semi-Definiteness

		γ			
		0.25	0.375	0.50	1.00
p=10	Extreme	0.97	0.92	0.88	0.76
	Illiquid	0.99	0.98	0.96	0.93
	Medium	1.00	1.00	1.00	0.97
	Liquid	1.00	1.00	1.00	1.00
p=50	Extreme	0.01	0.00	0.00	0.00
	Illiquid	0.00	0.00	0.00	0.00
	Medium	0.00	0.00	0.00	0.00
	Liquid	0.23	0.10	0.04	0.01

fnorm Results in dim=10

Legend:
RK
Block
RMT

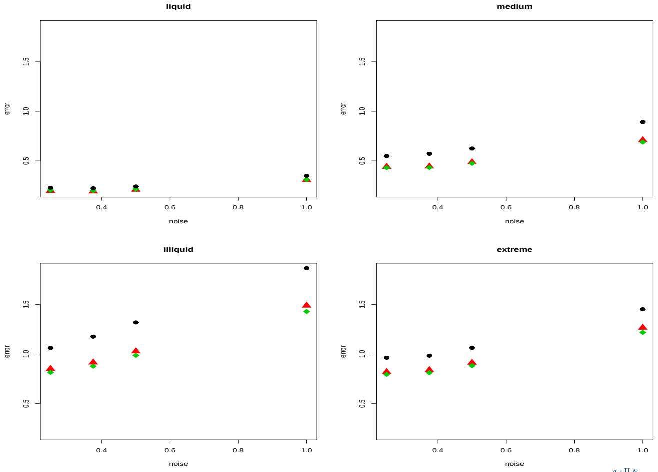
# block:
Liquid=5
Medium=5
Illiquid=5
Extreme=2



fnorm Results in dim=50

Legend:
RK
Block
RMT

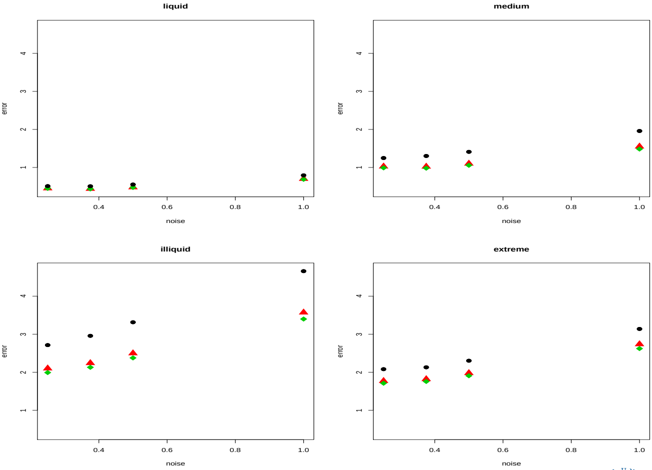
# block:
Liquid=5
Medium=5
Illiquid=5
Extreme=2



fnorm Results in dim=100

Legend:
RK
Block
RMT

# block:
Liquid=5
Medium=5
Illiquid=5
Extreme=2



Implications

- Blocking significantly increases the performance of the estimator
- RMT successfully removes negative or small eigenvalues
- Performance increases with the cross-sectional dimension
- Computationally very tractable even if the cross-sectional dimension is very high!

$\Rightarrow$ Pragmatic approach to handle huge covariances. Relevant in practice!